

1 Grundlagen

Das folgende Kapitel gibt einen kurzen Überblick über BPMN-Diagramme und das Mining Software Repositories, als Forschungsgebiet.

1.1 BPMN

Die Business Process Model and Notation (BPMN, deutsch Geschäftsprozessmodell und -notation) ist eine Spezifikation, die eine grafische Notation zum Ausdrücken von Geschäftsprozessen in einem Geschäftsprozessdiagramm (englisch Business Process Diagram, BPD) bereitstellt. Sie ist ein öffentlicher Standard, der von OMG und BPMI verwaltet wird. Das gesamte Dokument, das BPMN ausführlich beschreibt, umfasst mehr als fünf hundert Seiten und kann von der offiziellen Internetseite OMG kostenlos heruntergeladen werden. Siehe: [2].

Das Hauptziel von BPMN, wie in der Spezifikation steht, ist es, eine Notation bereitzustellen, die für alle Geschäftsbenutzer leicht verständlich ist, von den Business-Analysten, die die ersten Entwürfe der Prozesse erstellen, bis zu den technischen Entwicklern, die für die Implementierung der Technologie verantwortlich sind, die diese Prozesse durchführt, und schließlich auf die Geschäftsleute, die diese Prozesse verwalten und überwachen. Somit schafft BPMN eine standardisierte Brücke für die Lücke zwischen Geschäftsprozessdesign und Prozessimplementierung [8].

Grundsätzlich basiert BPMN auf der Überarbeitung anderer Notationen und Methoden, insbesondere des Aktivitätsdiagramms der Unified Modeling Language (UML), des UML-EDOC-Geschäftsprozesses, IDEF, ebXML BPSS, des Aktivitäts-Entscheidungsflussdiagramms, RosettaNet, LOVeM und Event Driven Process Chains [7]. Die aktuelle Version von BPMN (2.0.2) definiert unter anderem das XML-basierte Format für das Speichern und die Übertragung von BPMN-Diagrammen.

Insgesamt werden in der Spezifikation zur BPMN vier Arten von BPD definiert:

- Prozessdiagramm (Process Diagram)
- Kollaborationsdiagramm (Collaboration Diagram)
- Choreographie-Diagramm (Choreography Diagram)
- Konversationsdiagramm (Conversation Diagram)[8].

Für die Bildung von BPD stellt Spezifikation zur BPMN eine Menge von Elementen sowie Informationen für ihre Bedeutung und Regeln, wie sie kombiniert wer-

den können, bereit. Auf solche Weise werden Syntax und Semantik von BPMN-Diagrammen geregelt [4]. Jedes Element gehört einer der fünf Kernelementkategorien [1]. Diese sind:

1. **Flow Objects** (Knoten)
2. **Data** (Daten)
3. **Connecting Objects** (Verbindende Objekte)
4. **Swimlanes** (Schwimmbahnen, die Teilnehmer darstellen)
5. **Artifacts** (Artefakte)

In der Tabelle 1.1 werden die wichtigsten Elemente von BPMN-Diagrammen dargestellt.

Tabelle 1.1: Grundlegende Modellierungselemente von BPMN-Diagrammen[1], [4]

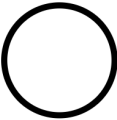

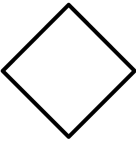
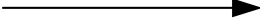
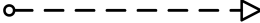
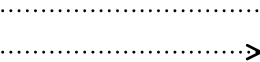
| Element | Beschreibung | Grafische Notation |
|---------------------------------|---|---|
| Event (Ereignis) | Ereignis markieren Zeitpunkte. Es gibt drei Arten von Ereignissen, je nachdem, wann sie den Fluss beeinflussen: Start, Zwischenzeit und Ende. |  |
| Activity (Aktivität) | Eine Aktivität ist ein Oberbegriff für die Arbeit, die ein Unternehmen in einem Prozess ausführt. Eine Aktivität kann atomar oder nicht atomar sein. |  |
| Gateway (Entscheidungspunkt) | Ein Entscheidungspunkt wird verwendet, um die Divergenz und Konvergenz von Sequenzflüssen zu steuern. Es gibt drei Arten von Entscheidungspunkten: XOR-Gateway, auch exklusives Gateway lässt genau einen Prozesspfad zu; UND-Gateway, paralleles Gateway (alle Pfade müssen durchschritten werden); OR-Gateway, inklusives Gateway (es muss mindestens ein Pfad gewählt werden). |  |
| Sequence Flow (Sequenzfluss) | Ein Sequenzfluss wird verwendet, um die Reihenfolge anzuzeigen, in der Aktivitäten ausgeführt werden. |  |
| Message Flow (Nachrichtenfluss) | Ein Nachrichtenfluss wird verwendet, um den Nachrichtenfluss zwischen zwei Teilnehmern anzuzeigen, die bereit sind, sie zu senden und zu empfangen. |  |
| Association (Assoziation) | Eine Assoziation wird verwendet, um Informationen und Artefakte mit BPMN-Grafikelementen zu verknüpfen. |  |

Tabelle 1.1: Grundlegende Modellierungselemente von BPMN-Diagrammen

| Element | Beschreibung | Grafische Notation |
|--|---|--------------------|
| Pool (Schwimm- becken) | Ein Pool ist die grafische Darstellung eines Teilnehmers an einer Kollaboration. Es fungiert auch als Swimlane und als grafischer Container, um eine Gruppe von Aktivitäten von anderen Pools zu trennen. Ein Pool kann einen Prozess referenzieren, muss aber nicht. D.h. ein Pool kann auch als eine Blackbox sein. | |
| Lane (Spur) | Eine Spur ist eine Unterpartition innerhalb eines Prozesses, manchmal innerhalb eines Pools, und erstreckt sich über die gesamte Länge des Prozesses, entweder vertikal oder horizontal. | |
| Data Object (Datenob- jekt) | Datenobjekte bieten Informationen dazu, welche Aktivitäten ausgeführt werden müssen und / oder was sie produzieren. Datenobjekte können ein einzelnes Objekt oder eine Sammlung von Objekten darstellen. | |
| Message (Nachricht) | Eine Nachricht wird verwendet, um den Inhalt einer Kommunikation zwischen zwei Teilnehmern darzustellen. | |
| Group (Gruppe) | Eine Gruppe ist eine Menge grafischer Elemente, die sich in derselben Kategorie befinden. Gruppen bieten die Möglichkeit, Kategorien von Objekten visuell im Diagramm anzuzeigen. | |
| Text An- notation (Textanmer- kung) | Textanmerkungen sind ein Mechanismus für einen Modellierer, um dem Leser eines BPMN-Diagramms zusätzliche Textinformationen bereitzustellen. | |

Wir möchten uns in der Abbildung 1.1 ein einfaches BPMN-Prozessdiagramm anschauen.

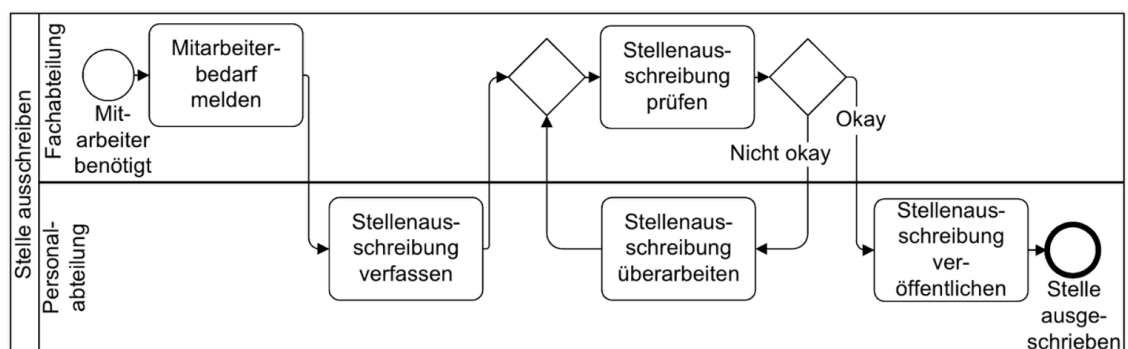


Abbildung 1.1: Ein einfaches BPMN-Prozessdiagramm. Übernommen aus [3].

Der Prozess „Stelle ausschreiben“, an dem Fachabteilung und Personalabteilung beteiligt sind, beginnt wenn ein neuer Mitarbeiter benötigt wird. Wenn das der Fall ist, meldet die Fachabteilung den Mitarbeiterbedarf. Nachdem das gemacht wird, verfasst die Personalabteilung eine Stellenausschreibung, die von Fachabteilung überprüft wird. Falls sie nicht okay ist, wird sie von Personalabteilung überarbeitet und wieder von Fachabteilung überprüft, wobei sich dieser Vorgang mehrmals wiederholen kann. Wenn Stellenausschreibung okay ist, wird sie von Personalabteilung veröffentlicht. Somit ist der Prozess „Stelle ausschreiben“ zu Ende [3].

Jetzt versuchen wir die Abbildung 1.1 nach Mal zu beschreiben und dabei die dargestellte Elemente zu nennen. Insgesamt besteht dieses Diagramm aus folgenden Elementen:

- Pool
- Lane
- Start Event
- Activity
- Sequence Flow
- XOR-Gateway
- End Event

Das gesamte Prozess befindet sich innerhalb des Pools und hat den Namen „Stelle ausschreiben“. Das Pool ist in zwei Bahnen (eng. „Lanes“) geteilt. Die Teilung in Bahnen macht es deutlich, welche Aktivitäten von Fachabteilung und welche von Personalabteilung gemacht werden. Der Prozess startet mit Start Event. Die vom Start Event ausgehende Kante (Sequence Flow) zeigt, welches Element als Nächstes durchlaufen wird. So kommen wir zur Aktivität „Mitarbeiterbedarf melden“, der die Aktivität „Stellenausschreibung verfassen“ folgt. Nach letzter passieren wir einen XOR-Gateway, der nur eine ausgehende Kante (Sequence Flow) hat. Somit erreichen wir die Aktivität „Stellenausschreibung prüfen“. Dieser Aktivität folgt das zweite XOR-Gateway mit zwei ausgehenden Kanten. Jetzt kann es wegen der Exklusivität der XOR-Gateway nur in eine Richtung weiter gehen. Die Richtung der Sequenz wird durch die Bedingungen „Okay“ oder „Nicht Okay“ bestimmt. Somit kann es wieder zur Aktivität „Stellenausschreibung prüfen“ und der beschriebenen Vorgang wiederholt sich, oder zur Aktivität „Stellenausschreibung veröffentlichen“. Nach dieser Aktivität wird das letzte Element, nämlich End Event, erreicht, was das Ende des Prozesses bedeutet.

1.2 Mining Software Repositories

Das Mining Software Repositories (MSR) -Forschungsgebiet analysiert und verbindet die umfangreichen, in den Softwareprojekten verfügbaren Daten, um interessante und umsetzbare Informationen zu Softwaresystemen und -projekten zu finden [5].

Beispiele für Software-Repositories sind:

- **Historische Repositories** wie Quellcodeverwaltungs-Repositories, Fehler-Repositories und archivierte Kommunikationen zeichnen verschiedene Informationen über die Entwicklung und den Fortschritt eines Projekts auf [5].
- **Laufzeit-Repositories** wie Bereitstellungsprotokolle enthalten Informationen zur Ausführung und Verwendung einer Anwendung an einer oder mehreren Bereitstellungsstandorten [5].
- **Code-Repositories** wie Git-Versionsverwaltungssysteme GitHub und GitLab oder SourceForge. Solche Hosting-Plattformen enthalten den Quellcode verschiedener Anwendungen, die von mehreren Entwicklern entwickelt wurden.

Programmierer erstellen und sammeln während der Softwareentwicklung viele Daten, die alle automatisch abgerufen und analysiert werden können. Wie zum Beispiel:

- Der Quellcode des Programms. Dies ist die wichtigste Eingabe für die mögliche Analyse.
- Durch die während der Ausführung der Software gesammelte Daten können Profile erhalten werden, die mitteilen, welche Teile der Software häufig verwendet werden und welche nicht.
- Den Produkten liegen möglicherweise zusätzliche Dokumentationen bei, beispielsweise Konstruktionsdokumente oder Anforderungsdokumente. Diese können auch wichtige Funktionen enthalten, die erklären, warum Code so aussieht, wie er aussieht.
- Die resultierende Software kann statisch analysiert werden und bietet Funktionen wie Komplexitätsmetriken oder Abhängigkeiten.
- Versionsarchive zeichnen die am Produkt vorgenommenen Änderungen auf, einschließlich wer, wann, wo und warum. Versionsarchive können viel über die Historie eines Projekts aussagen, wenn die gespeicherten Änderungen alle logisch getrennt sind und die gespeicherten Rationalen systematisch und konsistent verwendet werden [6].

Der Algorithmus für das MSR kann wie folgt formuliert werden:

1. **Ziel des MSR bestimmen**, d.h. welche Daten sollen gesammelt werden. Diese Daten werden später für die Analyse verwendet.
2. **Datenabruf**. Um die ständige Verfügbarkeit der gezielten Daten sicherzustellen und einen schnellen Zugriff auf sie zu ermöglichen, sollten lokale Kopien erstellt werden.
3. **Datenkonvertierung** (optional). Data Mining erfordert, dass die Daten nicht nur heruntergeladen und verarbeitet werden, sondern auch, dass man viele ihrer Merkmale im Voraus versteht. Daher sollten die Daten den Voraussetzungen und Annahmen entsprechen, die vorher getroffen wurden.

4. **Datenextraktion.** Nachdem die Daten auf lokaler Festplatte in einer geeigneten Form gesichert wurden, können Sie verarbeitet werden. Die Verarbeitung umfasst das Extrahieren, Filtern und Speichern der Inhalte der Ressourcen in einem dauerhaften und für den Menschen lesbaren Format. Zum Beispiel die gefilterte Informationen können in einer relationalen Datenbankmanagementsystem gespeichert werden.
5. **Datenanalyse.** In diesem Schritt wird die gezielte Analyse der gespeicherten Daten durchgeführt [6].

Literaturverzeichnis

- [1] Business process modeling notation (bpmn). https://www.service-architecture.com/articles/web-services/business_process_modeling_notation_bpmn.html. Accessed: 2019-04-15.
- [2] formal/13-12-09. <https://www.omg.org/spec/BPMN/2.0.2>. Accessed: 2019-04-15.
- [3] Thomas Allweyer. *BPMN 2.0 - Business Process Model and Notation - Einführung in den Standard für die Geschäftsprozessmodellierung*. BoD â Books on Demand, Norderstedt, 1. aufl. edition, 2015.
- [4] Kocian Claudia. *Geschäftsprozessmodellierung mit BPMN 2.0 Business Process Model and Notation im Methodenvergleich*. HNU Working Paper Nr. 16. Neu-Ulm University of Applied Sciences, Wileystraße 1, D-89231 Neu-Ulm, 2011.
- [5] A. E. Hassan. The road ahead for mining software repositories. In *2008 Frontiers of Software Maintenance*, pages 48–57, Sep. 2008.
- [6] K. S. Herzig and A. Zeller. Mining your own evidence. In *Making Software - What Really Works, and Why We Believe It*, pages 517–529. O'Reilly Media, Inc., Sebastopol, 2011.
- [7] Vivek Kale. *Enterprise Process Management Systems - Engineering Process-Centric Enterprise Systems using BPMN 2.0*. CRC Press, Boca Raton, Fla, 2018.
- [8] OMG. *Business Process Model and Notation (BPMN)*. Needham USA(2013), 2013.